X2016 – MAP 311
PC 7 – 12 juin 2017 – Estimateurs, TCL, vecteurs gaussiens

Igor Kortchemski – igor.kortchemski@cmap.polytechnique.fr

Corrigé des exercices non traités sur http://www.normalesup.org/~kortchem/MAP311 *un peu après la PC.*

Exercice 1. **(Manipulations sur les gaussiennes)** On rappelle qu'une variable aléatoire gaussienne X de paramètres (m, σ^2) a pour fonction caractéristique $\phi(t) = \mathbb{E}\left[e^{itX}\right] = e^{imt - \frac{\sigma^2 t^2}{2}}$.

(1) Soit $X = \mathcal{N}(0, 1)$ une loi gaussienne centrée réduite. Quelle est la loi de $m + \sigma X$?

(2) Soient $X = \mathcal{N}(m_1, \sigma_1^2)$ et $Y = \mathcal{N}(m_2, \sigma_2^2)$ deux variables aléatoires gaussiennes indépendantes. Quelle est la loi de $X + Y$? Ce résultat reste-t-il vrai si X et Y ne sont pas indépendantes ?

(3) Soit $(X_k)_{k \geq 1}$ une suite de gaussiennes centrées réduites indépendantes. On pose

$$Y_n = \frac{1}{n} \sum_{k=1}^{n} \sqrt{k} X_k.$$

Étudier la convergence en loi de Y_n.

1 Estimateur du maximum de vraisemblance

Exercice 2. **(EMV pour des variables exponentielles)** On observe un échantillon $(X_1, \ldots, X_n)$ de n temps d'attente du RER B, qu'on suppose indépendants et de même loi exponentielle de paramètre $\theta > 0$ inconnu.

(1) Déterminer l'estimateur du maximum de vraisemblance $\widehat{\theta_n} = \theta_n(X_1, \ldots, X_n)$ de θ et montrer qu'il converge presque sûrement vers θ lorsque $n \to \infty$.

(2) Cet estimateur est-il sans biais ?
On rappelle que la densité de la loi $\Gamma(a, \lambda)$ *est* $\frac{1}{\Gamma(a)} \lambda^a x^{a-1} e^{-\lambda x} \mathbb{1}_{x>0}$.

(3) Démontrer que $\sqrt{n}(\widehat{\theta_n} - \theta)$ converge en loi vers une loi qu'on déterminera.
Indication. On pourra utiliser la « méthode delta ».

Exercice 3. **(Plus appliqué)** Une colonie de vampires a élu domicile dans un château des Carpates, et la comptesse D. souhaite estimer leur nombre. Pour cela, une nuit de pleine lune, la comptesse en capture 10, leur mord les oreilles, puis les relâche. La nuit suivante, elle en capture 10 au hasard. Trois ont une morsure aux oreilles.

Aidez la comptesse D. en utilisant un estimateur du maximum de vraisemblance.

Pour des questions, demande d'explications etc., n'hésitez pas à m'envoyer un mail.

2 Problème récapitulatif

Exercice 4. **(Modèle auto-régressif d'ordre 1 AR(1)).** Soit $(Y_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes et de même loi $\mathcal{N}(m,\sigma^2)$. Soient $a \in \mathbb{R}$ et X_0 une variable aléatoire de loi $\mathcal{N}(m_0,\sigma_0^2)$ indépendante de $(Y_n)_{n\geq 1}$. Pour tout $n \geq 1$, on définit la suite récurrente aléatoire $(X_n)_{n\geq 1}$ par

$$X_n = aX_{n-1} + Y_n.$$

Il s'agit d'un cas particulier de *séries temporelles*, utilisées pour modéliser l'évolution passée d'une quantité pour en prévoir le comportement futur.

Première partie.

(1) Montrer que $(X_0,\dots,X_n)$ est un vecteur gaussien.

(2) Déterminer la loi de X_n et exprimer $\mathrm{Cov}(X_k,X_n)$ en fonction de $\mathrm{Var}(X_k)$ pour $0 \leq k \leq n$. Trouver la valeur de c telle que $X_0 + cX_1$ soit indépendant de X_0.

(3) À quelle condition sur a la suite (X_n) converge-t-elle en loi ? Quelle est alors la loi limite ? Quelle est la loi de X_n si X_0 a cette loi limite ?

(4) Montrer que si $a \in]-1,1[$, le vecteur (X_n,X_{n+1}) converge en loi vers un vecteur gaussien dont on déterminera les paramètres.

(5) Pour $a \in]-1,1[$, étudier la convergence en loi de la moyenne empirique $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.

Deuxième partie. On suppose maintenant que $(Y_n)_{n\geq 1}$ est une suite de variables aléatoires indépendantes et de même loi $\mathcal{N}(0,\sigma^2)$ (avec σ connu), que $X_0 = 0$ et que a est inconnu.

(6) Soit g la densité d'une loi gaussienne $\mathcal{N}(0,\sigma^2)$. Montrer qu'une densité f_n de $(X_1,\dots,X_n)$ au point $(x_1,\dots,x_n)$ est

$$f_n(x_1,\dots,x_n) = g(x_1)g(x_2 - ax_1)\cdots g(x_n - ax_{n-1}).$$

(7) Déterminer l'estimateur du maximum de vraisemblance $\widehat{a_n} = \widehat{a_n}(X_1,\dots,X_n)$ de a.

3 À chercher pour la prochaine fois

Exercice 5. Soit $(X_k)_{k\geq 1}$ des variables aléatoires indépendantes de même loi uniforme sur $[0,\theta]$ (avec θ inconnu).

(1) Montrer que l'estimateur du maximum de vraisemblance $\widehat{\theta}_n = \widehat{\theta}_n(X_1,\dots,X_n)$ de θ est $\widehat{\theta}_n = \max(X_1,\dots,X_n)$.

(2) Montrer que $W_n = n\left(1 - \frac{\widehat{\theta}_n}{\theta}\right)$ converge en loi quand $n \to \infty$ et déterminer sa limite.

4 Pour aller plus loin

Exercice 6. **(Estimateurs linéaires)** Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes de même loi et de carré intégrable. Trouver l'estimateur $\widehat{\theta}_n$ de la moyenne $\theta = \mathbb{E}[X_1]$, qui soit sans biais (c'est-à-dire $\mathbb{E}\left[\widehat{\theta}_n\right] = \theta$ et de variance minimale dans la classe des estimateurs linéaires $\widehat{\theta}_n = \sum_{k=1}^n a_k X_k$.

Exercice 7. Soit $g : [0,1] \to [0,1]$ une fonction (mesurable) bornée. On souhaite calculer $m = \int_0^1 g(x)dx$. On pose $\sigma^2 = \int_0^1 g(x)^2 dx - m^2$. Soient X et Y des variables aléatoires indépendantes de loi uniforme sur $[0,1]$. On pose

$$U = \mathbb{1}_{Y \leq g(X)}, \qquad V = g(X), \qquad W = \frac{g(X) + g(1-X)}{2}.$$

(1) Calculer l'espérance et la variance de U, V, W. Comparer les variances de U et V.

(2) Proposer trois méthodes de type Monte-Carlo pour calculer m.

On suppose dans la suite que g est monotone.

(3) Vérifier que $\mathbb{E}[g(X)g(1-W)] \leq m^2$ et comparer les variances de V et W.

Indication. On pourra montrer que $(g(x)-g(y))(g(1-x)-g(1-y)) \leq 0$ pour tout $x, y \in [0,1]$.

Soit $(X_i)_{i\geq 1}$ une suite de variables aléatoires indépendantes de loi uniforme sur $[0,1]$.

(4) On considère les estimateurs suivants de m :

$$A_n = \frac{1}{2n}\sum_{k=1}^{2n} g(X_k) \qquad \text{et} \qquad \frac{1}{2n}\sum_{k=1}^{n}(g(X_k) + g(1-X_k)).$$

Montrer qu'ils sont sans biais. Lequel possède la plus petite variance ?

(5) Dans le cas où $g(x) = x^2$, déterminer le nombre n de simulations nécessaires garantissant une précision relative de 1% sur le calcul de m en erreur quadratique avec A_n et B_n (la précision relative étant $\frac{\mathsf{Var}(A_n)}{m^2}, \frac{\mathsf{Var}(B_n)}{m^2}$).